(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 May 2005 (19.05.2005)

PCT

(10) International Publication Number
WO 2005/045016 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number:
PCT/US2004/026194

(22) International Filing Date: 11 August 2004 (11.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/494,300	11 August 2003 (11.08.2003)	US
60/545,657	18 February 2004 (18.02.2004)	US

(71) Applicant *(for all designated States except US)*: **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, CA 94063 (US).

(72) Inventors: **KREBBER, Anke**; 3500 Louis Road, Palo Alto, CA 94303 (US). **NEWMAN, Lisa Marie**; 1002 King Street, Redwood City, CA 94061 (US). **DAVIS, S. Christopher**; 104 Winfield Street, San Francisco, CA 94110 (US). **JENNE, Stephane J.**; 821 El Camino Real #202, Burlingame, CA 94010 (US).

(74) Agent: **POCHOPIEN, Donald J.**; MCANDREWS, HELD & MALLOY, LTD., 500 W. Madison Street, Suite 3400, Chicago, IL 60661 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

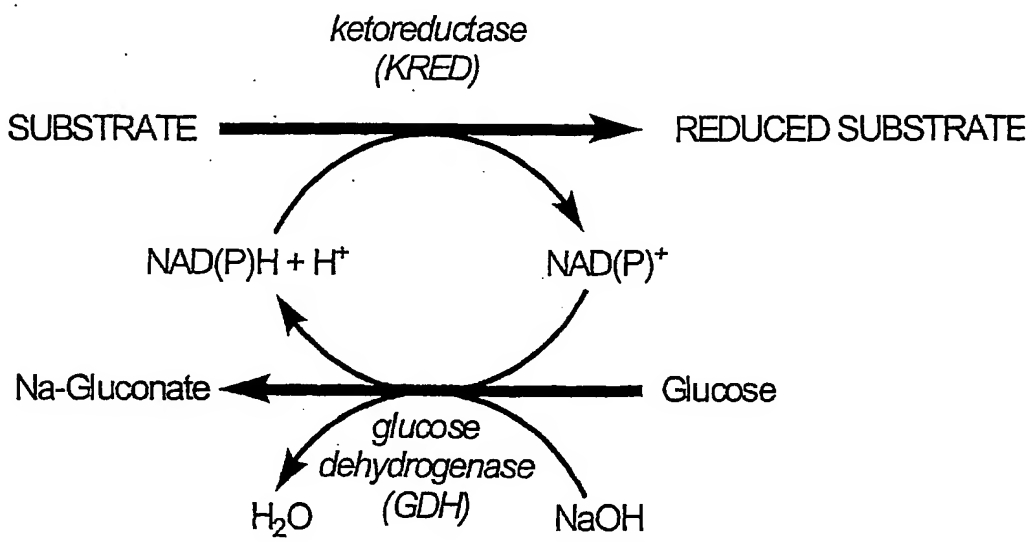
Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMPROVED GLUCOSE DEHYDROGENASE POLYPEPTIDES AND RELATED POLYNUCLEOTIDES



(57) Abstract: The present invention is directed to glucose dehydrogenase (GDH) polypeptides that have enhanced GDH activity and/or thermostability relative to the backbone wild-type glucose dehydrogenase polypeptide. In addition, the present invention is directed to a plynucleotide that encodes for the GDH polypeptides of the present invention, to nucleic acid sequences comrpising the polynucleotides, to expression vectors comprising the polynucleotides operatively linked to a promoter, to host cells transformed to express the GDH polypeptides, and to a method for producing the GDH polypeptides of the present invention.

WO 2005/045016 A2

IMPROVED GLUCOSE DEHYDROGENASE POLYPEPTIDES AND RELATED POLYNUCLEOTIDES

FIELD OF THE INVENTION

[01] The present invention is related to the field of enzymology, and particularly to the field of glucose dehydrogenase enzymology. More specifically, the present invention is directed to glucose dehydrogenase polypeptides having improved enzymatic activity (*i.e.*, high substrate turnover) and stability, and to polynucleotides sequences encoding for the improved glucose dehydrogenase polypeptides. The present invention is useful because the glucose dehydrogenase polypeptides can be coupled to oxido- or reductase enzymes to produce synthetic organic chemicals or precursors in high yields.

BACKGROUND OF THE INVENTION

[02] Glucose dehydrogenase [EC1.1.1.47] or "GDH" catalyzes the conversion of β-glucose and nicotinamide adenine dinucleotide (NAD) to gluconolactone and reduced nicotinamide adenine dinucleotide (NADH). NAD serves as a co-factor in this reaction and may be phosphorylated in the form of NADP. GDH is an important enzyme for use in clinical tests and the food industry. GDH is also applied as a catalyst for chemical conversions where it serves a role in the regeneration of NADH and NADPH in enzymatic carbonyl reductions, such as aldehydes and ketones.

[03] *Bacillus* species have been an excellent source of GDH. The enzyme from *B. megaterium* M1286 was purified to homogeneity and found to be a homotetramer of 30,000 DA subunits with pH optimum of 8.0-9.0 depending on buffer conditions and uses either nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) as cofactor (Pauly H.E. and Pfleiderer G., Hoppe Seylers Z. Physiol. Chem. 1975 356:1613-23). The enzyme from *Cryptococcus uniguttulatus* Y 0033 has a pH optimum of 6.0-8.0, an optimum temperature of 55° C and a molecular weight of 110 kDa (U.S. Pat. 4,877,733). The enzyme from *Pseudomonas* sp. FH1227 has a pH optimum of

8.5-9.0, an optimum temperature of 55° C and a molecular weight of 101 kDa (U.S. Pat. 5,298,411).

[04] Commercially applied GDHs are primarily derived from microorganisms. Initially, GDH was produced by fermentation of the natural host organisms such as *B. megaterium* ATCC 39118 (U.S. Pat. 4,542,098), *Bacillus cereus* DSM 1644 (US4397952), *Cryptococcus uniguttulatus* Y 0033 (U.S. Pat. 4,877,733) and *Pseudomonas* sp. FH1227 (U.S. Pat. 5,298,411). Since then, GDH encoding genes have been identified, cloned and expressed in heterologous hosts such as *Escherichia coli.*

[05] The *Bacillus subtilis* 61297 GDH gene was expressed in *E. coli* and exhibited the same physicochemical properties as the enzyme produced in its native host (Vasantha et al. Proc. Natl. Acad. Sci. USA 1983 80:785). The gene sequence of the *B. subtilis* GDH gene was reported by Lampel, K. A., Uratani, B., Chaudhry, R., Ramaley, R. F., and Rudikoff S., "Characterization of the developmentally regulated *Bacillus subtilis* glucose dehydrogenase gene," *J. Bacteriol.* 166, 238-243 (1986) and Yamane,K., Kumano,M. and Kurita,K., "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome: determination of the sequence of a 146 kb segment and identification of 113 genes," *Microbiology* 142 (Pt 11), 3047-3056 (1996), and is found in Genbank under Accession Nos. M12276 and D50453.

[06] Similarly, gene sequences were determined for GDH from *B. cereus* ATCC14579 (Nature 2003 423:87-91; Genbank Acc. No. AE017013) and *B. megaterium* (Eur. J. Biochem. 1988 174:485-490, Genbank Acc. No. X12370; J. Ferment. Bioeng. 1990 70:363-369, Genbank Acc. No. D90044). The GDH enzymes from *B. subtilis* and *B. megaterium* are approximately 85% homologous (J. Theor. Biol. 1986 120:489-497).

[07] It has been well established that GDH enzymes suffer from limited stability. Ramaley and Vasantha reported that presence of glycerol in extraction and purification buffers is absolutely necessary to retain activity for GDH from *B. subtilis* (J. Biol. Chem. 1983 258:12558-12565). The enzyme instability can be largely attributed to the dissociation of the tetramer into its

monomers, which is an equilibrium process that is controlled by environmental factors such as pH and ionic strength (Maurer and Pfleiderer, Z. Naturforsch. 1987 42: 907-915). This has lead to the isolation and studies of GDH from other *Bacillus* sp. such as *B. megaterium*. For instance, U.S. Pats. 5,114,853 and 5,126,256 and Baik et al. Appl. Microbiol. Biotechnol. 2003 61:329-335 describe GDH encoding genes from *B. megaterium* and mutants thereof that exhibit increased thermostability and that can be produced in recombinant *E. coli* hosts. However, there remains an industrial need for GDH enzymes that not only have increased thermostability but that also have enhanced enzymatic activity. The above referenced publications and patents, and all other publications and patents referenced herein, are hereby incorporated by reference herein in their entirety.

BRIEF SUMMARY OF THE INVENTION

[08] The present invention has multiple aspects. In one aspect, the present invention is directed to a polypeptide having at least 1.5 times, typically 1.5 to about 25 times, more typically from 1.5 to about 11 times, the GDH activity of the wild-type GDH of SEQ ID NO: 2 (such as determined by the method of Example 4) and being selected from the group consisting of:

(a) a polypeptide having an amino acid sequence which has at least 91% homology, preferably at least 95% homology, and more preferably at least 98% homology with the amino acid sequence of SEQ ID NO: 54, 74, 84, 160, 164 or 168 (hereinafter "homologous polypeptides");

(b) a polypeptide encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73, 83, 159, 163 or 167; (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii) (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.);

(c) a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 comprising a substitution, deletion, and/or insertion of one to six amino acids;

(d) a fragment of (a), (b) or (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2, ; and

(e) a polypeptide of (a), (b) or (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7. In one embodiment, the present invention is also directed to a variant GDH polypeptide as described herein in isolated and purified form. In another embodiment, the present invention is directed to a variant GDH polypeptide as described herein in lyophilized form. In yet another embodiment, the present invention is directed to a composition comprising a variant GDH polypeptide as described herein and a suitable carrier, typically a buffer solution, more typically a buffer solution having a pH between 6.0 and 8.0.

[09] The novel GDH polypeptides of the present invention have enhanced GDH activity (>1.5 fold) relative to the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2 and typically vary from SEQ ID NO: 2 by 1-7 amino acid residues, more typically by 1-6 amino acid residues, even more typically by 1-5 amino acid residues, and most typically by 1-4 amino acid residues. For purposes of the present invention, the degree of homology between two amino acid sequences was determined using the Needleman Wunsch global alignment algorithm, i.e., using dynamic programming algorithm for Global Alignment Scoring Matrix: PAM 120 matrix with gap penalties for introducing gap = -22.183 and extending gap = -1.396. The percent identity = number of identical residues between the first sequence and the second sequence divided by the length of first sequence in alignment (with gaps)(p) indicates partial match. *See* Needleman, S.B. & Wunsch, C.D., "A general method applicable to the search for similarities in the amino acid sequence of two proteins," Journal of Molecular Biology, 48:443-453 (1970).

[10] The various residue positions of the *B. subtilis* GDH polypeptide that have been substituted to yield enhanced GDH activity and/or thermostability are summarized in Table 1 herein. The amino acid sequences for a number of the inventive GDH polypeptides that have demonstrated enhanced GDH activity and/or thermostability at 50° C are disclosed herein as SEQ ID NOS: 6, 8, 10,

12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166 and 168. The polynucleotide sequences encoding for the above described inventive GDH polypeptides have SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 75, 77, 83, 85, 87, 89, 91, 93, 95, 97, 99, 99, 101, 103, 105, 107, 109, 11, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165 and 167, respectively. Another GDH polypeptide (SEQ ID NO: 52) of the present invention, which is encoded by the polynucleotide of SEQ ID NO: 51, is based upon the amino acid sequence of GDH (SEQ ID NO: 4) from *B megaterium* and differs therefrom by six amino acid residues.

[11] In a preferred embodiment, the present invention is directed to the novel glucose dehydrogenase polypeptides of SEQ ID NOS: 54, 74, 84, 160, 164, and 168 that have enhanced glucose dehydrogenase activity and/or enhanced thermostability relative to the wild-type glucose dehydrogenase of SEQ ID NO: 2.

[12] In yet another embodiment, the present invention is directed to polypeptide having glucose dehydrogenase enzyme activity and being selected from the group consisting of a GDH polypeptide having at least 84% sequence identity with SEQ ID NO: 52, a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 72, and a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 58.

[13] In its second aspect, the present invention is directed to a polynucleotide sequence that encodes for the correspondingly referenced GDH polypeptide. Given the degeneracy of the genetic code, the present invention is also directed to any polynucleotide that encodes for the above referenced GDH polypeptides of the present invention. In another preferred embodiment, the present invention is directed to certain specific polynucleotides of SEQ ID NOS: 53, 73,

83, 159, 163, and 167 that encode for the novel glucose dehydrogenase polypeptides of SEQ ID NOS: 54, 74, 84, 160, 164 and 168, respectively.

[14] In a third aspect, the present invention is directed to a nucleic acid construct, a vector, or a host cell comprising a polynucleotide sequence encoding a GDH polypeptide of the present invention operatively linked to a promoter.

[15] In a fourth aspect, the present invention is directed to a method of making a GDH polypeptide of the present invention comprising (a) cultivating a host cell comprising a nucleic acid construct comprising a nucleic acid sequence encoding a GDH polypeptide of the present invention under conditions suitable for production of the polypeptide; and (b) recovering the polypeptide.

BRIEF DESCRIPTION OF SEVERAL VIEWS OF THE DRAWINGS

[16] FIG. 1 exemplifies an oxidation-reduction cycle wherein glucose is oxidized by GDH to gluconic acid in the presence of NAD^+ (or $NADP^+$) to produce the corresponding reduced form NADH (or NADPH), respectively, which in turn drives the reduction of a substrate to a reduced substrate while being oxidized back to NAD (or NADP) by a reductase. The gluconic acid formed in this reaction is neutralized by sodium hydroxide to sodium-gluconate.

[17] FIGS. 2A-2B in combination provide a table comparing the % amino acid identity of the GDH polypeptides of the present invention versus the GDH polypeptides of the indicated prior art references. In col. 4 of FIG. 2, the GDH polypeptide of *B subtilis* (S06-3) has the same amino acid sequence as disclosed in EP 955375 (col. 8). To generate FIGS 2A-2B, alignments were done using dynamic programming algorithm for Global Alignment Scoring Matrix: PAM 120 matrix with gap penalties for introducing gap = -22.183 and extending gap = -1.396. The percent identity = number of identical residues between the first sequence and the second sequence divided by the length of first sequence in alignment (with gaps)(p) indicates partial match. *See* Needleman, S.B. & Wunsch, C.D., "A general method applicable to the search for similarities in the

amino acid sequence of two proteins," Journal of Molecular Biology, 48:443-453 (1970).

[18] FIG. 3 is a 4036 bp expression vector (pCK110900) of the present invention comprising a P15A origin of replication (P15A ori), a lacI repressor, a CAP binding site, a lac promoter (lac), a T7 ribosomal binding site (T7g10 RBS), and a chloramphenicol resistance gene (camR).

[19] The foregoing summary, as well as the following detailed description of certain embodiments of the present invention, will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there is shown in the drawings, certain embodiments. It should be understood, however, that the present invention is not limited to the arrangements and instrumentality shown in the attached drawings.

DETAILED DESCRIPTION OF THE INVENTION

[20] The present invention has multiple aspects. In one aspect, the present invention is directed to a polypeptide having at least 1.5 times, typically 1.5 to about 25 times, more typically from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2 (such as determined by the method of Example 4) and being selected from the group consisting of:

(a) a polypeptide having an amino acid sequence which has at least 91% homology, preferably at least 95% homology, and more preferably at least 98% homology with the amino acid sequence of SEQ ID NO: 54, 74, 84, 160, 164 or 168 (hereinafter "homologous polypeptides");

(b) a polypeptide encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73 or 83, (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii) (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.);

(c) a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 comprising a substitution, deletion, and/or insertion of one to six amino acids;

(d) a fragment of (a), (b) or (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2; and

(e) a polypeptide of (a), (b) or (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

[21] Unless otherwise noted, as used throughout this specification, the terms "percent identity," "% identity," "percent identical," and "% identical" are used interchangeably herein to refer to the percent amino acid sequence identity that is determined using the Needleman Wunsch global alignment algorithm, i.e., using dynamic programming algorithm for Global Alignment Scoring Matrix: PAM 120 matrix with gap penalties for introducing gap = -22.183 and extending gap = -1.396. The percent identity = number of identical residues between the first sequence and the second sequence divided by the length of first sequence in alignment (with gaps)(p) indicates partial match. *See* Needleman, S.B. & Wunsch, C.D., "A general method applicable to the search for similarities in the amino acid sequence of two proteins," Journal of Molecular Biology, 48:443-453 (1970).

[22] As used herein, the terms "glucose dehydrogenase" and "GDH" are used interchangeably herein to refer to a polypeptide that has the ability to catalyze the conversion of glucose and nicotinamide adenine dinucleotide (NAD) to gluconolactone and reduced nicotinamide adenine dinucleotide (NADH). Alternatively, the phosphorylated cofactors NADP and NADPH can replace NAD and NADH in the above reaction. In nature, GDH is made up of four subunits that are loosely held together in a homo-tetramer. Based upon the crystal structure of wild-type *B. megaterium* GDH polypeptide (SEQ ID NO: 4) (Yamamoto et al. J. Biochem. 2001 129:303-312), residue positions 188-217 of the polypeptide define a protein loop region that is involved in NAD^+ and glucose binding.

[23] In use, the enhanced GDH polypeptides of the present invention are preferably coupled to a synthetic reaction as a cofactor regeneration system (See Figure 1) to provide a continuing source of reduced cofactor. As used herein, the term "cofactor" refers to a non-protein compound that operates in

combination with an enzyme that catalyzes a reaction of interest. Suitable cofactors employed with the GDH polypeptides of the present invention include NADP (nicotinamide-adenine dinucleotide phosphate) and NAD (nicotinamide adenine dinucleotide).

[24] The term "cofactor regeneration system" refers herein to a set of reactants that participate in a reaction that regenerates a utilized cofactor back to its pre-reaction state. An example is the regeneration of oxidized cofactor regeneration back to reduced cofactor, e.g., NADP to NADPH. The reduced (regenerated) cofactor is then capable of participating in a reaction with a substrate and an enzyme, such as a reducing enzyme, to produce the reduced substrate and the oxidized (utilized) cofactor, which can again be regenerated by the cofactor regeneration system. The above-described operation of the glucose/glucose dehydrogenase cofactor regeneration system is exemplified in Figure 1.

[25] In FIG. 1, the reaction catalyzed by the reducing enzyme is shown as being coupled to the glucose dehydrogenase cofactor regeneration system. The term "coupled" is used herein to refer to the use of the reduced form of cofactor in the reduction of a substrate, and the concomitant use of the oxidized form of the same cofactor, generated in the aforementioned reaction, in the oxidation of a component (e.g., glucose) of the cofactor regeneration system, which generates the reduced form of the same cofactor. One possible limiting factor in the overall reaction speed in a coupled system is the speed (activity) of the GDH polypeptide in regenerating cofactor.

[26] The GDH polypeptides of the present invention have enhanced GDH activity (such as measured by the method of Example 4) that is 1.5 fold to about 11 fold greater than the GDH activity of the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2, and typically vary from SEQ ID NO: 2 by 1-7 amino acid residues, more typically by 1-6 amino acid residues, even more typically by 1-5 amino acid residues, and most typically by 1-4 amino acid residues. Preferably, the GDH polypeptides of the present invention have enhanced GDH activity that is 2.5 fold to about 11 fold greater than the GDH

activity of the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2. More preferably, the GDH polypeptides of the present invention have enhanced thermostability after heat treatment at 50° C for 20 minutes that is 1.5 fold to about 15 fold greater than the GDH activity of the backbone GDH polypeptide from *B. subtilis* of SEQ ID NO: 2. Thermostability is determined by measuring the residual GDH activity (such as by the method of Example 4) remaining after heat treatment of the GDH polypeptide at 50° C for 20 minutes.

[27] The amino acid sequences for a number of the inventive GDH polypeptides that have demonstrated enhanced GDH activity and/or thermostability at 50° C are disclosed herein as SEQ ID NOS: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166 and 168. The polynucleotide sequences encoding for the above described inventive GDH polypeptides have SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 75, 77, 83, 85, 87, 89, 91, 93, 95, 97, 99, 99, 101, 103, 105, 107, 109, 11, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165 and 167, respectively.

[28] In yet another aspect, the present invention is directed to GDH polypeptides that have enhanced activity in coupled reactions.

[29] In another embodiment, the present invention is directed to a GDH polypeptide encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73, 83, 159, 163 or 167 (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii) (J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.). For polynucleotides of at least 100 nucleotides in length, low to very high stringency conditions are defined as

prehybridization and hybridization at 42° C in 5x SSPE, 0.3% SDS, 200 mμg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures.

[30] For polynucleotides of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2x SSC, 0.2% SDS at least at 50° C (low stringency), at least at 55° C (medium stringency), at least at 60° C. (medium-high stringency), at least at 65° C (high stringency), and at least at 70° C. (very high stringency).

[31] In another embodiment, the present invention is directed to a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 having a substitution, deletion, and/or insertion of one to six amino acids therefrom, and having from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2, such as determined by the method of Example 4. Preferably, amino acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to six amino acids; small amino- or carboxyl-terminal extensions; a small linker peptide; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

[32] Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine, proline, cysteine and methionine). Amino acid substitutions, which do not generally alter the specific activity are known in the art and are described, for example, by H. Neurath and R. L. Hill, 1979, In, The Proteins, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val,

Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly as well as these in reverse.

[33] In another embodiment, the present invention is directed to a fragment of (a), (b) or (c), as described above in the first paragraph of the Detailed Description, that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2, such as determined by the method of Example 4. By the term "fragment" is meant that the polypeptide has a deletion of 1 to 10 amino acid residues from the carboxy terminus, the amino terminus, or both. Preferably, the deletion is 1 to 10 residues from the carboxy terminus, more preferably, the deletion is 1 to 5 residues from the carboxy terminus.

[34] In yet another embodiment, the present invention is directed to a GDH polypeptide of (a), (b) or (c), as described above in the first paragraph of the Detailed Description, that retains more than 80% of the initial (pre-incubation) GDH activity after 20 minutes of incubation at 50° C and pH 7. Preferably, the polypeptides of the invention retain at least 85% of the initial residual activity, more preferably at least 90% residual activity after 20 minutes incubation at 50° C and pH 7. The initial GDH activity is readily determined by an assay for GDH activity, such as described in Example 4 herein.

[35] Another GDH polypeptide (SEQ ID NO: 52) of the present invention is based upon the amino acid sequence of the GDH polypeptide (SEQ ID NO: 4) from *B megaterium* and differs therefrom by six amino acid residues. This GDH polypeptide is encoded by the polynucleotide of SEQ ID NO: 51.

[36] In yet another embodiment, the present invention is directed to polypeptide having glucose dehydrogenase enzyme activity and being selected from the group consisting of a GDH polypeptide having at least 84% sequence identity with SEQ ID NO: 52, a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 72, and a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 58. These percent identities were obtained by ClustalW analysis (version W 1.8 available form European Bioinformatics Institute, Cambridge, UK), counting the number of identical matches in the alignment and dividing such number of identical matches by the length of the

reference sequence, and using the following default ClustalW parameters to achieve slow/accurate pairwise optimal alignments – Gap Open Penalty:10; Gap Extension Penalty:0.10; Protein weight matrix: Gonnet series; DNA weight matrix: IUB; Toggle Slow/Fast pairwise alignments = SLOW or FULL Alignment.

Polynucleotides

[37] In its second aspect, the present invention is directed to a polynucleotide sequence that encodes for a GDH polypeptide of the present invention. Given the degeneracy of the genetic code, the present invention is also directed to any polynucleotide that encodes for the above referenced GDH polypeptides of the present invention. In a preferred embodiment, the present invention is directed to certain specific polynucleotides of SEQ ID NOS: 53, 73, 83, 159, 163 and 167 that encode for the novel glucose dehydrogenase polypeptides of SEQ ID NOS: 54, 74, 84, 160, 164 and 168, respectively.

[38] To make the improved GDH polypeptides of the present invention, one starts with one or more wild-type polynucleotides that encode a GDH polypeptide. The term "wild-type" polynucleotide means that the nucleic acid fragment does not comprise any mutations from the form isolated from nature. The term "wild-type" protein means that the protein will be active at a level of activity found in nature and typically will comprise the amino acid sequence as found in nature. Thus, the term "wild type" or "parental sequence" indicates a starting or reference sequence prior to a manipulation of the invention.

[39] Suitable sources of wild-type GDH as a starting material to be improved are readily identified by screening genomic libraries for the GDH activities described herein. *See e.g.*, Example 4. A particularly suitable source of GDH is the *Bacillus sp.* bacteria as found in nature. *See* Example 1. Using the published glucose dehydrogenase gene sequences for *B. subtilis* (*e.g.*, EP 955375) and *B. megaterium* (*e.g.*, U.S. Pat. 5,114,853), primers for amplification of the genes from their respective gene libraries were created using conventional techniques and have the following sequences:

B. subtilis forward primer (SEQ ID NO: 80):

5'-GAATTCGCCCATATGTATCCGGATTTAAAAGG-3'

B. subtilis reverse primer (SEQ ID NO: 81):

5'-TGGCCGGATCCTCATTAACCGCGGCCTGCCTGGA-3'

B. megaterium forward primer (SEQ ID NO: 82):

5'-GAATTCGCCCATATGTATAAAGATTTAGAAGG-3'

B. megaterium reverse primer (SEQ ID NO 83):

5'-GGCCGGATCCTCATTATCCGCGTCCTGCTTGGA-3'

[40] Using a forward and reverse primer pair in a conventional polymerase chain reaction (PCR) to amplify the appropriate portion of the gene from *B. subtilis* and *B. megaterium*, several PCR products were obtained. Each PCR product was cloned into an expression vector and operatively linked behind a *lac* promoter. Upon screening the clones for GDH activity (*e.g.*, by the assay of Example 4), several clones were found to express active GDH and these genes were sequenced. A first DNA sequence (SEQ ID NO: 1) designated as S06-3 and encoding a GDH polypeptide identical to the published *Bacillus subtilis* GDH (SEQ ID NO: 2) was obtained from *Bacillus subtilis*. A second DNA sequence (SEQ ID NO: 3), designated as M02-6 and encoding a GDH polypeptide that was 98.5% identical to the published *Bacillus megaterium* GDH (SEQ ID NO: 4), was obtained from *Bacillus megaterium*. These DNA sequences were utilized as the starting material for developing the improved polypeptides and polynucleotides of the present invention.

[41] In addition to the PCR primers described above, other PCR primers of different lengths could be used as well. See, e.g., Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Other conventional nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) could also be used.

[42] Once a suitable starting material has been identified, a non-naturally occurring and mutated and/or evolved enzyme, having unknown glucose dehydrogenase activity is readily generated using any one of the well-known

mutagenesis or directed evolution methods. See, e.g., Ling, et al., "Approaches to DNA mutagenesis: an overview," Anal. Biochem., 254(2):157-78 (1997); Dale, et al., "Oligonucleotide-directed random mutagenesis using the phosphorothioate method," Methods Mol. Biol., 57:369-74 (1996); Smith, "In vitro mutagenesis," Ann. Rev. Genet., 19:423-462 (1985); Botstein, et al., "Strategies and applications of in vitro mutagenesis," Science, 229:1193-1201 (1985); Carter, "Site-directed mutagenesis," Biochem. J., 237:1-7 (1986); Kramer, et al., "Point Mismatch Repair," Cell, 38:879-887 (1984); Wells, et al., "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites," Gene, 34:315-323 (1985); Minshull, et al., "Protein evolution by molecular breeding," Current Opinion in Chemical Biology, 3:284-290 (1999); Christians, et al., "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling," Nature Biotechnology, 17:259-264 (1999); Crameri, et al., "DNA shuffling of a family of genes from diverse species accelerates directed evolution," Nature, 391:288-291; Crameri, et al., "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology, 15:436-438 (1997); Zhang, et al., "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening," Proceedings of the National Academy of Sciences, U.S.A., 94:45-4-4509; Crameri, et al., "Improved green fluorescent protein by molecular evolution using DNA shuffling," Nature Biotechnology< 14:315-319 (1996); Stemmer, "Rapid evolution of a protein in vitro by DNA shuffling," Nature, 370:389-391 (1994); Stemmer, "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution," Proceedings of the National Academy of Sciences, U.S.A., 91:10747-10751 (1994); WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746 which issued to Arnold, *et al.* on March 25, 2003 and is entitled "Method for creating polynucleotide and polypeptide sequences."

[43] Any of these methods can be applied to generate GDH polynucleotides. To maximize any diversity, several of the above-described techniques can be used sequentially. Typically, a library of variant polynucleotides is created by one mutagenic or evolutionary technique and their expression products are

screened to find the polypeptides having the highest GDH activity. Then, a second mutagenic or evolutionary technique is applied to polynucleotides encoding the most active polypeptides to create a second library, which in turn is screened for GDH activity by the same technique. The process of mutating and screening can be repeated as many times as needed, including the insertion of point mutations, to arrive at a polynucleotide that encodes a polypeptide with the desired activity, thermostability, and cofactor preference.

[44] Alternatively, polynucleotides and oligonucleotides of the invention can be prepared by standard solid-phase methods, according to known synthetic methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined (e.g., by enzymatic or chemical litigation methods, or polymerase mediated methods) to form essentially any desired continuous sequence. For example, polynucleotides and oligonucleotides of the invention can be prepared by chemical synthesis using, *e.g.*, the classical phosphoramidite method described by *Beaucage et al.* (1981) *Tetrahedron Letters* 22:1859-69, or the method described by *Matthes et al.* (1984) EMBO J. 3:801-05, *e.g.*, as it is typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized, *e.g.*, in an automatic DNA synthesizer, purified, annealed, ligated and cloned in appropriate vectors.

[45] In addition, essentially any nucleic acid can be custom ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company, Midland, TX, The Great American Gene Company (Ramona, CA), ExpressGen Inc., Chicago, IL, Operon Technologies Inc. (Alameda, CA), and many others. Similarly, peptides and antibodies can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, Inc. (http://www.htibio.com), BMA Biomedicals Ltd. (U.K.), Bio.Synthesis, Inc., and many others.

[46] Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., *Carruthers et al., Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and *Adams et al., J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained

either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

[47] General texts which describe molecular biological techniques useful herein, including mutagenesis, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, volume 152 Academic Press, Inc., San Diego, CA ("Berger"); *Sambrook et al.*, Molecular Cloning - A Laboratory Manual (2nd Ed.), volumes 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook"); and Current Protocols in Molecular Biology, F.M. *Ausubel et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (supplemented through 2000) ("Ausubel")). Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qβ-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA) are found in Berger, Sambrook, and Ausubel, as well as *Mullis et al.*, (1987) U.S. Patent No. 4,683,202; PCR Protocols A Guided to Methods and Applications (*Innis et al.*, eds.) Academic Press Inc. San Diego, CA (1990); Arnheim & Levinson (October 1, 1990) Chemical and Engineering News 36-47; The Journal Of NIH Research (1991) 3:81-94; *Kwoh et al.* (1989) Proc. Natl. Acad. Sci. USA 86:1173; *Guatelli et al.* (1990) Proc. Natl. Acad. Sci. USA 87:1874; *Lomell et al.* (1989) J. Clin. Chem. 35:1826; *Landegren et al.*, (1988) Science 241:1077-1080; Van Brunt (1990) Biotechnology 8:291-294; Wu and Wallace, (1989) Gene 4:560; *Barringer et al.* (1990) Gene 89:117, and Sooknanan and Malek (1995) Biotechnology 13:563-564. Improved methods of cloning *in vitro* amplified nucleic acids are described in *Wallace et al.*, U.S. Pat. No. 5,426,039. Improved methods of amplifying large nucleic acids by PCR are summarized in *Cheng et al.* (1994) Nature 369:684-685 and the references therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using

reverse transcriptase and a polymerase. *See,* Ausubel, Sambrook and Berger, *all supra.*

[48] It will be appreciated by those skilled in the art due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding GDH polypeptides of the invention may be produced, some of which bear substantial identity to the nucleic acid sequences explicitly disclosed herein.

[49] In the present case, several round No. 1 libraries were created by applying a variety of mutagenic techniques to the coding region of the *B. subtilis gdh* gene (SEQ ID NO: 1) or to the coding region of the *B. megaterium gdh* gene (SEQ ID NO: 3), as obtained by PCR.

[50] To obtain expression of the variant gene encoding a GDH, the variant gene was first operatively linked to one or more heterologous regulatory sequences that control gene expression to create a nucleic acid construct, such as an expression vector or expression cassette. Thereafter, the resulting nucleic acid construct, such as an expression vector or expression cassette, was inserted into an appropriate host cell for ultimate expression of the GDH polypeptide encoded by the shuffled gene. A "nucleic acid construct" is defined herein as a nucleic acid molecule, either single-or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid combined and juxtaposed in a manner that would not otherwise exist in nature. Thus, in one aspect, the present invention is directed to a nucleic acid construct comprising a polynucleotide encoding a GDH polypeptide of the present invention.

[51] The term "nucleic acid construct" is synonymous with the term "expression cassette" when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" is defined herein as a nucleic acid sequence, which directly specifies the amino acid sequence of its protein product. The boundaries of a genomic coding sequence are generally determined by a ribosome binding site (prokaryotes) or by the ATG start codon (eukaryotes) located just upstream of the open reading frame at the 5' end of the mRNA and a transcription

terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

[52] An isolated polynucleotide encoding a GDH polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[53] The term "control sequence" is defined herein to include all components, which are necessary or advantageous for the expression of a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[54] The term "operably linked" is defined herein as a configuration in which a control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence directs the expression of a polypeptide.

[55] The control sequence may be an appropriate promoter sequence. The "promoter sequence" is a relatively short nucleic acid sequence that is recognized by a host cell for expression of the longer coding region that follows. The promoter sequence contains transcriptional control sequences, which mediate the expression of the polypeptide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained

from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[56] For bacterial host cells, suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, include the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), as well as the tac promoter (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242: 74-94; and in Sambrook et al., 1989, supra.

[57] For filamentous fungal host cells, suitable promoters for directing the transcription of the nucleic acid constructs of the present invention include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof.

[58] In a yeast host, useful promoters are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and

Saccharomyces cerevisiae 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos et al., 1992, Yeast 8:423-488.

[59] The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator, which is functional in the host cell of choice, may be used in the present invention.

[60] Preferred terminators for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

[61] Preferred terminators for yeast host cells are obtained from the genes for Saccharomyces cerevisiae enolase, Saccharomyces cerevisiae cytochrome C (CYC1), and Saccharomyces cerevisiae glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos et al., 1992, supra.

[62] The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention. Preferred leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[63] The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and

which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence that is functional in the host cell of choice may be used in the present invention. Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.

[64] The control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. The foreign signal peptide coding region may be required where the coding sequence does not naturally contain a signal peptide coding region.

[65] Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to enhance secretion of the polypeptide. However, any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

[66] Effective signal peptide coding regions for bacterial host cells are the signal peptide coding regions obtained from the genes for Bacillus NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are described by Simonen and Palva, 1993, Microbiological Reviews 57: 109-137.

[67] Effective signal peptide coding regions for filamentous fungal host cells are the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase.

[68] Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos et al., 1992, supra.

[69] The control sequence may also be a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (WO 95/33836).

[70] Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[71] It may also be desirable to add regulatory sequences, which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include the lac, tac, and trp operator systems. In yeast host cells, suitable regulatory systems include the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include the TAKA

alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[72] Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene, which is amplified in the presence of methotrexate, and the metallothionein genes, which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the GDH polypeptide of the present invention would be operably linked with the regulatory sequence.

Expression Vectors

[73] In another aspect, the present invention is also directed to a recombinant expression vector comprising a polynucleotide of the present invention (which encodes a GDH polypeptide of the present invention), and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[74] The recombinant expression vector may be any vector (e.g., a plasmid or virus), which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

[75] The expression vector may be an autonomously replicating vector, *i.e.*, a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

[76] The expression vector of the present invention preferably contains one or more selectable markers, which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol (Example 1) or tetracycline resistance. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3.

[77] Selectable markers for use in a filamentous fungal host cell include, but are not limited to, amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate decarboxylase), sC (sulfate adenyltransferase), and trpC (anthranilate synthase), as well as equivalents thereof. Preferred for use in an Aspergillus cell are the amdS and pyrG genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the bar gene of *Streptomyces hygroscopicus*.

[78] The vectors of the present invention preferably contain an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome. For integration into the host cell genome, the vector may rely on the nucleic acid sequence

encoding the polypeptide or any other element of the vector for integration of the vector into the genome by homologous or nonhomologous recombination.

[79] Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

[80] For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are P15A, the origins of replication of plasmids pBR322, pUC19, pACYCl77, which has the P15A origin of replication), or pACYC184 which permit replication in E. coli; and pUB110, pE194, pTA1060, or pAM.beta.1 which permit replication in Bacillus. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of replication may be one having a mutation which makes it's functioning temperature-sensitive in the host cell (see, e.g., Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75: 1433).

[81] More than one copy of a nucleic acid sequence of the present invention may be inserted into the host cell to increase production of the gene product. An increase in the copy number of the nucleic acid sequence can be obtained by

integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the nucleic acid sequence where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the nucleic acid sequence, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

[82] The procedures used to ligate the elements described above to construct the recombinant nucleic acid construct and expression vectors of the present invention are well known to one skilled in the art (see, *e.g.*, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, N.Y.).

[83] Many of the expression vectors for use in the present invention are commercially available. Suitable commercial expression vectors include p3xFLAGTM™ expression vectors from Sigma-Aldrich Chemicals, St. Louis MO., which includes a CMV promoter and hGH polyadenylation site for expression in mammalian host cells and a pBR322 origin of replication and ampicillin resistance markers for amplification in *E. coli*. Other suitable expression vectors are pBluescriptII SK(-) and pBK-CMV, which are commercially available from Stratagene, LaJolla CA, and plasmids that are derived from pBR322 (Gibco BRL), pUC (Gibco BRL), pREP4, pCEP4 (Invitrogene) or pPoly (Lathe et al., 1987, Gene 57, 193-201).

Host Cells

[84] Host cells for use in expressing the expression vectors of the present invention include but are not limited to, bacterial cells, such as *E. coli*, Streptomyces and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* or *Pichia pastoris* (ATCC Accession No. 201178)); insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are well known in the art.

[85] By way of example, *Escherichia coli* W3110 was transformed by an expression vector for expressing the shuffled genes of the present invention. The expression vector was created by operatively linking a variant gene of the present invention to the *lac* promoter under control of the *lacI* repressor gene. The expression vector also contained the P15A origin of replication and the chloroamphenicol resistance gene. The transformed *Escherichia coli* W3110 was cultured under appropriate culture medium containing chloramphenicol such that only transformed *E coli* cells that expressed the expression vector survived. *See e.g.*, Example 1.

Purification

[86] Once the GDH polypeptides were expressed by the variant genes in *E. coli*, the polypeptides are purified from the cells and or the culture medium using any one or more of the well known techniques for protein purification, including lysozyme treatment, sonication, filtration, salting, ultra-centrifugation, affinity chromatography, and the like. Suitable solutions for high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available under the trade name CelLytic B™ from Sigma-Aldrich of St. Louis MO. A suitable process for purifying GDH polypeptides sufficiently from cell lysate for applications in a chemical process is disclosed in Example 3 herein.

Screening

[87] Screening clones of the GDH polypeptides from the expression libraries for enhanced GDH activity is typically performed using the standard biochemistry technique of monitoring the rate of formation of NADH or NADPH, via an increase in absorbance or fluorescence. Such a procedure is described in Example 4 herein.

[88] After screening the libraries after the first round of mutations, an improved GDH polypeptide, which had the mutation I165T relative to the *B. subtilis* GDH backbone of SEQ ID NO: 2, provided a 2.6 fold increase in initial GDH activity relative to the wild-type *B. subtilis* GDH (SEQ ID NO: 2). Thereafter, additional rounds of directed evolution were performed and the

resulting exemplary GDH polypeptides of the present invention are listed in Table 1 below along with their mutations and activity relative to the wild-type (w-t) *B. subtilis* GDH backbone of SEQ ID NO: 2:

Table 1

GDH Peptide No.	Mutations	X-fold Initial GDH Activity over w-t GDH from *B. subtilis*	Increased Thermostability @ 50° C, 20 min
SEQ ID NO: 84	I165M, V209A, I242V	***	No
SEQ ID NO: 52	V209A	*	No
SEQ ID NO: 54	I165M, P194T	**	Yes
SEQ ID NO: 56	I165T	**	Yes
SEQ ID NO: 64	I165M, V209A, I242V, Q252L	**	Yes
SEQ ID NO: 74	I165L	**	Yes
SEQ ID NO: 160	I165M, P194T, A197K, K204E, K206R	***	Yes
SEQ ID NO: 164	I165 M, E170K, P194T, A197K, K204E, K206R, E222D, S237C	***	Yes
SEQ ID NO: 168	A16T, I165M, P194T, A197K, K204E, K206R	***	Yes

* 1.5-2.4 fold increase
** 2.5-3.4 fold increase
*** 3.5-5.5 fold increase

[89] Comparing the GDH polypeptides of SEQ ID NOS: 64 and 84 in Table 1 above, it is seen that the Q252L mutation conferred heat stability on the GDH polypeptide, while reducing some initial GDH activity.

[90] The GDH polypeptide of SEQ ID NO: 74, which has the single change I165L relative to the *B. subtilis* GDH backbone of SEQ ID NO: 2, provides a 5 fold increase in initial GDH activity, and a 13 fold increase in activity in a coupled chemistry process relative to the wild-type *B. subtilis* GDH (SEQ ID NO: 2).

[91] More preferred GDH polypeptides of the present invention are those polypeptides having 95% homology, more preferably 97% homology, and even more preferred 100% homology with the polypeptides of SEQ ID NOs: 160, 164 and 168. As shown above, each of these polypeptides has from 5-8 mutations but has the following five mutations in common: I165M, P194T, A197K, K204E and K206R. Thus, stated in other terms, one embodiment of the present invention is directed to a GDH polypeptide of SEQ ID NO: 2 having from 5-8 residue substitutions wherein five of the residue substitution are I165M, P194T, A197K, K204E, and K206R.

[92] Only a very few ($\leq$ 0.5%) of the mutations to the wild-type *B. subtilis* GDH (SEQ ID NO: 2) backbone were found to be beneficial. Specifically, for every 1000 clones screened, there occurred only 3-5 single point or double point mutations that were beneficial. In fact, many of the mutations were found to be detrimental. For example, Y253C rendered the GDH polypeptide inactive, whereas Q252L slightly reduced the initial GDH activity wild-type *B. subtilis* GDH (SEQ ID NO: 2). Interestingly, the beneficial effects of one mutation were not found to be additive with the beneficial effects of another mutation. Thus, for example, it was discovered that the combination of a first mutation that increased GDH activity 2 fold compared to the wild-type activity, with a second mutation at a second residue position that increased GDH activity 3 fold compared to the wild-type activity most often did not result in a GDH polypeptide that had a 5 or 6 fold increase in GDH activity.

[93] The GDH polypeptides of the present invention have the activities described herein, as well as other desirable properties, *e.g.*, altered temperature and/or pH optimums, solvent resistance (*e.g.*, butyl acetate), and the like. Moreover, the GDH polynucleotide may be mutated or evolved to generate libraries that can be screened to identify those modified GDH polypeptides having the ability to preferentially accept other compounds as cofactors, such as, for example, NADP (also referred to as $NADP^+$).

[94] The polynucleotides encoding the GDH polypeptides of the present invention may be codon optimized for optimal production from the host

organism selected for expression. Those having ordinary skill in the art will recognize that tables and other references providing codon preference information for a wide range of organisms are readily available. See *e.g.*, Henaut and Danchin, "*Escherichia coli* and *Salmonella*," Neidhardt, et al. Eds., ASM Press, Washington D.C., p. 2047-2066 (1996).

[95] Screening for transformed cells that express GDH is, in general, a two-step process. First, one physically separates the cells and then determines which cells do and do not possess a desired property. Selection is a form of screening in which identification and physical separation are achieved simultaneously by expression of a selection marker, which, in some genetic circumstances, allows cells expressing the marker to survive while other cells die (or vice versa). Exemplary screening markers include luciferase, β-galactosidase, and green fluorescent protein. Selection markers include drug and toxin resistance genes, such as resistance to chloramphenicol, ampicillin and the like. Although spontaneous selection can and does occur in the course of natural evolution, in the present methods selection is performed by man.

[96] The GDH polynucleotides generated by the mutagenesis or directed evolution method are screened in accordance with the protocol described in Example 4 to identify those having enhanced activity that are suitable for inclusion as an improved GDH polypeptide of the present invention.

[97] The following sequence summarizes the diversity of the GDH polypeptides of the present invention relative to the wild type *B. subtilis* GDH polypeptide of SEQ ID NO: 2, as also disclosed in EP 955375, wherein one or more of the amino acid residues designated as "X" followed by the residue number are replaced to the GDH polypeptides of the present invention:

[98]

M Y P D L K G K V V X_{11} I T G A X_{16} S G L G K A M A I R F G K E Q A K V V I N Y Y S N K Q D P X_{46} X_{47} V K E E V I X_{54} A G G E A V V V X_{63} G D V T X_{68} E E D V K N X_{75} V Q T A I K E F G T L D I M I N X_{92} A G X_{95} X_{96} N P V P S H E M P L K D W D K V I X_{114} T N L T G A F L G S R E A I K Y F V E N D I K G N V I N M S S V H E V I P W P L F V H Y A A S K G G X_{165} K L M T X_{170} T L A L E Y A X_{178} K G I R V N N I G P G A I N T X_{194} I N X_{197} X_{198} K X_{200} X_{201} X_{202} X_{203} X_{204} X_{205} X_{206} A D X_{209} X_{210} X_{211} X_{212} I P M

X_{216} Y I X_{219} X_{220} P X_{222} E I A A V A A W L A S X_{234} E A X_{237} Y V T G X_{242} T L F A D G G M T X_{252} X_{253} P S F Q A G R G

The diversity of changes at various residue positions for the GDH polypeptides of the present invention are shown to the right of the arrow in Table 2 below and relative amino acid residues of wild-type GDH of SEQ ID NO: 2 (EP 055375) which are shown to the left of the arrow:

Table 2

X_{11}: A→ D,
X_{16}: A→ T
X_{46}: N→D
X_{47}: E→ D, K,
X_{54}: K→R
X_{63}: Q→ R
X_{68}: K→ N,
X_{75}: I→ , V
X_{92}: N→ S
X_{95}: L→ F
X_{96}: E→ A
X_{114}: G→ S
X_{165}: I→ M, L, V, T
X_{170} : E→K
X_{178}: P→ Q
X_{194}: P→ T
X_{197}: A→ K
X_{198}: E→ A, D, G, K, L
X_{200}: F→ M, Y
X_{201}: A→ G, L, S, T
X_{202}: D→ N
X_{203}: P→ D, N, R, S
X_{204}: K→ D, E, L, N, T, Q
X_{205}: Q→ N
X_{206}: K→ R
X_{209}: V→ A
X_{210}: E→ A, K
X_{211}: S→ G, T
X_{212}: M→ K
X_{216}: G→ L
X_{219}: G→ A
X_{220}: E→ Q
X_{222} : E→D
X_{234}: K→ E, S
X_{237}: S→C
X_{242}: I→ V

X_{252}: Q→L
X_{253}: Y→C

Example 1: Construction of Expression Constructs for Expression of Glucose Dehydrogenase

The genes for the glucose dehydrogenase were amplified using the polymerase chain reaction (PCR) from genomic DNA preparations of *Bacillus subtilis* and *Bacillus megaterium*. The primers for the amplification reactions were designed using the published *B. subtilis* and *B. megaterium* glucose dehydrogenase gene sequences, and were as follows:

B. subtilis forward primer (SEQ ID NO: 79):

5'-GAATTCGCCCATATGTATCCGGATTTAAAAGG-3'

B. subtilis reverse primer (SEQ ID NO: 80):

5'-TGGCCGGATCCTCATTAACCGCGGCCTGCCTGGA-3'

B. megaterium forward primer (SEQ ID NO: 81):

5'-GAATTCGCCCATATGTATAAAGATTTAGAAGG-3'

B. megaterium reverse primer (SEQ ID NO 82):

5'-GGCCGGATCCTCATTATCCGCGTCCTGCTTGGA-3'

[99] The PCR products were cloned into an expression vector of FIG. 3 behind a lac promoter under control of the lacI repressor gene, creating plasmids pGDHS06 or pGDHM02. The expression vector contained the P15A origin of replication (P15A ori) and the chloroamphenicol resistances gene (camR). Several clones were found to express active GDH and these genes were sequenced to confirm their sequences (see SEQ ID NOS: 1 (Glucose dehydrogenase S06-3) and 3 (Glucose dehydrogenase M02-6)).

Example 2: Production of GDH

[100] In an aerated agitated fermentor, 10.0L of growth medium containing 0.528g/L ammonium sulphate, 7.5g/L of di-potassium hydrogen phosphate trihydrate, 3.7g/L of potassium dihydrogen phosphate, 2g/L of Tastone-154 yeast extract, 0.05g/L ferrous sulphate, and 3ml/L of a trace element solution containing 2g/L of calcium chloride dihydrate, 2.2g/L of zinc sulfate septahydrate, 0.5g/L manganese sulfate monohydrate, 1g/L cuprous sulfate

heptahydrate, 0.1g/L sodium borate decahydrate and 0.5g/L EDTA, was brought to a temperature of 30° C.

[101] The fermentor was inoculated with a late exponential culture of *Escherichia coli* W3110 (pGDHS06 or pGDHM02) grown in a shake flask containing LB, 1% glucose (Sigma Chemical Co., St. Louis, MO), and 30 μg/ml chloroamphenicol (Sigma Chemical Co., St. Louis, MO) to a starting optical density at 600 nm (OD_{600}) of 0.5 to 2.0. The fermentor was agitated at 500-1500rpm and air was supplied to the fermentation vessel at 1.0-15.0 L/min, and the pH of the culture was controlled at 7.0 by addition of 20% v/v ammonium hydroxide. After the culture reached an OD_{600} of 40, the temperature was reduced to 25° C and the expression of glucose dehydrogenase was induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) (Sigma Chemical Corp., St. Louis, MO) to a final concentration of 1mM. The culture was grown for another 15 hours. After the induction, the cells were harvested by centrifugation and washed with 10mM potassium phosphate buffer, pH 7.0. The cell paste was used directly in the downstream recovery process or was stored at –80° C until use.

Example 3: GDH Enzyme Preparation

[102] The cell paste was washed by suspending 1 volume wet weight of cell paste in 3 volumes of 100mM Tris /sulfate (pH 7.2) followed by centrifugation at 5000g for 40 minutes in a Sorval 12BP. The washed cell paste was suspended in 2 volumes of 100mM Tris/sulfate (pH 7.2). The intracellular GDH was released from the cells by passing the suspension through a homogenizer in two passes using a pressure of 14,000 psig for the first pass and 8,000 psig for the second pass. The homogenate was centrifuged at 10,000 rpm in a Beckman lab centrifuge for 60 minutes. The supernatant was decanted and dispensed in shallow containers, frozen at –20° C and lyophilized.

Example 4: GDH Enzyme Activity Assay

[101] Cells were grown overnight at 30C in 2xYT with 0.5% glucose and 30ug/ml chloramphenicol. This culture was then diluted 20-fold into fresh LB

containing 30 ug/ml chloramphenicol and after 2 hours of growth at 37°C, 1mM IPTG (isopropyl thiogalactoside) was added. The culture (0.3 ml) was allowed to grow another 4-5 hours at 37°C.

[102] Lysis buffer contains 100 mM triethanolamine buffer (pH 7.0), 2 mg/ml PMBS (polymixin B sulfate), 1 mg/ml lysozyme, 1 mM PMSF (phenyl methyl sulfonyl fluoride).

[103] Cells were pelleted via centrifugation and lysed in 0.2 ml lysis buffer by shaking at room temperature for 1.5 hours.

[104] An aqueous assay mix was made that was 100 mM triethanolamine buffer (pH 7.0), 0.1 to 0.2 mM NADPH or NADH, and 100mM glucose. A solvent mixture consisting of 1 part ethyl-4-chloro-acetoacetate (ECAA) and 2 parts butyl acetate was added to the assay mixture to form a reaction mixture. The ratio of solvent mixture to assay mixture was 1:2, respectively. To test thermostability, the undiluted lysate was heated at 50°C and then added to the reaction mixture. The reaction was initiated by adding the diluted glucose dehydrogenate enzyme as a predissolved solution in 100 mM triethanolamine buffer (pH 7.0). The course of reaction was followed by measurement of the increase of absorbance at 340 nm or by the fluorescent emission of light at 440 nm as a function of time. The results were plotted as Absorbance units or relative fluorescent units (RFU) (NADPH or NADH) vs. time, and the slope of the plot determined (Absorbance units/min or RFU/min).

Example 5 KRED/GDH Coupled Chemistry Assay

[103] To a 100 mL vessel equipped with a pH electrode-controlled automatic titrator was charged a solution of glucose (7.5 g) in 100 mM triethanolamine pH 7 buffer (25 mL). To this solution were charged the two enzymes (100 mg KRED; 50 mg GDH) and NADP (6.25 mg). ("KRED" is ketoreductase or carbonyl reductase class (EC1.1.1.184) and is useful for the synthesis of optically active alcohols from the corresponding prochiral ketone substrate. Butyl acetate (10 ml) was then charged. Finally, ethyl 4-chloroacetoacetate (6 g) in butyl acetate (10 mL) was charged to the vessel. 4M NaOH is added

dropwise on demand by the automatic titrator (a pH of 6.85 was set as a lower limit) to constantly adjust the pH to 7.0. The reaction was complete when no more caustic was needed. The reaction rates were determined by measuring the amount of base added per unit time or by taking samples of the reaction mixture, extracting the sample 3 times with an equal volume of ethyl acetate, and analyzing the combined organic layers by gas chromatography to determine the amount of ethyl-S- 4-chloro-3-hydroxybutyrate produced per unit time.

[104] While the invention has been described with reference to certain embodiments, it will be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from its scope. Therefore, it is intended that the invention not be limited to the particular embodiment disclosed, but that the invention will include all embodiments falling within the scope of the appended claims.

CLAIMS

WHAT IS CLAIMED IS:

1. A polypeptide having at least 1.5 times the GDH activity of the wild-type GDH of SEQ ID NO: 2 and being selected from the group consisting of:

(a) a polypeptide having an amino acid sequence which has at least 91% homology with the amino acid sequence of SEQ ID NO: 54, 74, 84, 160, 164 or 168;

(b) a polypeptide encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73, 83, 159, 163 or 167 (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii) ;

(c) a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 comprising a substitution, deletion, and/or insertion of one to six amino acids;

(d) a fragment of (a), (b) or (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2; and

(e) a polypeptide of (a), (b) or (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

2. A polynucleotide encoding a glucose dehydrogenase polypeptide of claim 1.

3. An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes a polypeptide of claim 1.

4. An expression vector comprising a polynucleotide of claim 2 operatively linked to a promoter.

5. A host cell transformed to express a polynucleotide of claim 2.

6. A method of making a GDH polypeptide of claim 1, comprising (a) cultivating a host cell comprising a nucleic acid construct comprising a nucleic acid sequence encoding the polypeptide under conditions suitable for production of the polypeptide; and (b) recovering the polypeptide.

7. An isolated and purified GDH polypeptide of claim 1.

8. A GDH polypeptide of claim 1 in lyophilized form.

9. The polypeptide of claim 1, having an amino acid sequence which has at least 91% homology with the amino acid sequence of SEQ ID NO: 54, 74, 84, 160, 164 or 168.

10. The polypeptide of claim 1, encoded by a nucleic acid sequence which hybridizes under medium stringency conditions with either (i) the nucleotide sequence of SEQ ID NO: 53, 73, 83, 159, 163 or 167 (ii) a subsequence of (i) of at least 100 nucleotides, or (iii) a complementary strand of (i) or (ii).

11. The polypeptide of claim 1, being a variant of the polypeptide of SEQ ID NO: 54, 74, 84, 160, 164 or 168 comprising a substitution, deletion, and/or insertion of one to six amino acids.

12. The polypeptide of claim 1, being a fragment of (a), (b) or (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2.

13. The polypeptide of claim 12, being a fragment of (a) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2.

14. The polypeptide of claim 12, being a fragment of (b) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2.

15. The polypeptide of claim 1, being a fragment of (c) that has from 1.5 to about 11 times the GDH activity of the wild-type GDH of SEQ ID NO: 2.

16. The polypeptide of claim 1, being a polypeptide of (a), (b) or (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

17. The polypeptide of claim 16, being a polypeptide of (a) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

18. The polypeptide of claim 16, being a polypeptide of (b) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

19. The polypeptide of claim 16, being a polypeptide of (c) that retains more than 80% of the initial GDH activity after 20 minutes of incubation at 50° C and pH 7.

20. A composition comprising a polypeptide of claim 1 in a buffered medium.

21. A polypeptide having glucose dehydrogenase enzyme activity and being selected from the group consisting of a GDH polypeptide having at least 84% sequence identity with SEQ ID NO: 52, a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 72, and a GDH polypeptide having at least 98% sequence identity with SEQ ID NO: 58.

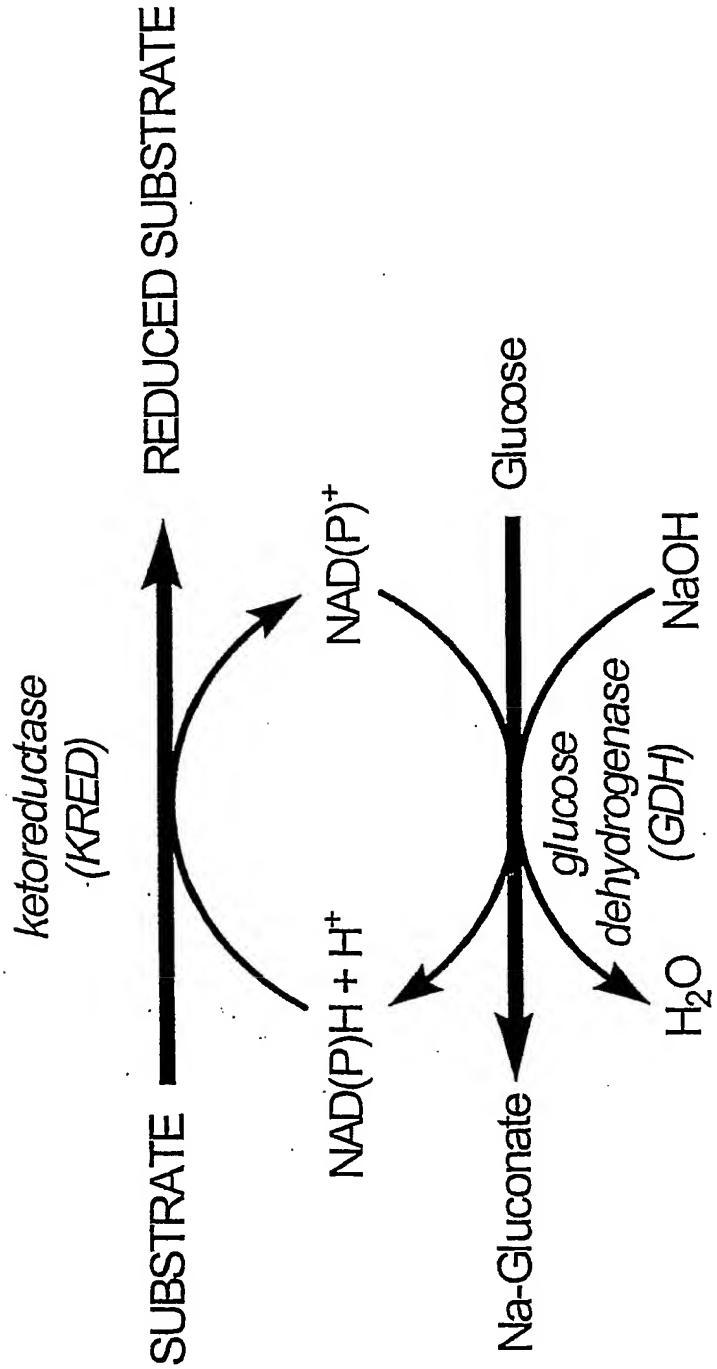

REDUCED SUBSTRATE
ketoreductase
(KRED)
NAD(P)+
Glucose
NaOH
glucose
dehydrogenase
(GDH)
H2O
NAD(P)H + H+
SUBSTRATE
Na-Gluconate

FIG. 1

SEQ ID NO	WO 200049039	US 5114853	S06-3	JP 04258293	JP 04258289	JP 02072878	EP955375	EP285949 DE3711881	EP1213354	EP1013758	DE3931716	AAA22463 (NCBI No)
78	82.8	82.0	99.2	81.6	82.0	82.0	99.2	82.0	83.1	99.2	80.8	98.1
76	82.8	82.0	98.9	81.6	82.0	82.0	98.9	82.0	83.1	98.9	80.8	97.7
74	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.8	98.5
72	82.0	81.2	98.1	80.8	81.2	81.2	98.1	81.2	82.4	98.1	80.1	96.9
70	82.4	82.0	99.2	81.6	81.6	82.0	99.2	81.6	83.1	99.2	80.5	98.1
68	82.4	82.0	98.9	81.6	81.6	82.0	98.9	81.6	83.1	98.9	80.5	97.7
66	82.0	81.2	98.1	80.8	81.2	81.2	98.1	81.2	82.4	98.1	80.5	96.9
64	82.4	81.6	98.5	81.2	81.6	81.6	98.5	81.6	82.8	98.5	80.5	97.3
62	82.0	81.6	98.5	81.2	81.2	81.6	98.5	81.2	82.8	98.5	80.5	97.3
60	81.6	81.2	98.1	80.8	80.8	81.2	98.1	80.8	82.4	98.1	79.7	96.9
58	81.6	81.2	98.5	80.8	80.8	81.2	98.5	80.8	82.4	98.5	79.7	97.3
56	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.8	98.5
54	82.8	82.4	99.2	82.0	82.0	82.4	99.2	82.0	83.5	99.2	80.8	98.1
52	82.8	97.3	83.5	96.9	82.0	97.3	83.5	82.0	99.2	83.5	95.4	82.8
50	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.8	98.5
48	82.8	82.0	99.6	81.6	82.0	82.0	99.6	82.0	83.1	99.6	80.8	98.5
46	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.8	98.5
44	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5
40	82.8	82.4	99.2	82.0	82.0	82.4	99.2	82.0	83.5	99.2	80.8	98.1
38	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5

FIG. 2A

SEQ ID NO	WO 200049039	US 5114853	S06-3	JP 04258293	JP 04258289	JP 02072878	EP955375	EP285949 DE3711881	EP1213354	EP1013758	DE3931716	AAA22463 (NCBI No)
36	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.8	98.5
34	82.8	82.4	99.2	82.0	82.0	82.4	99.2	82.0	83.5	99.2	80.8	98.1
32	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5
30	82.8	82.4	99.2	82.0	82.0	82.4	99.2	82.0	83.5	99.2	80.8	98.1
28	82.8	82.4	99.2	82.0	82.0	82.4	99.2	82.0	83.5	99.2	80.8	98.1
26	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.8	98.5
24	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5
22	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.5	98.5
20	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5
18	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.5	98.5
16	82.4	82.0	99.2	81.6	81.6	82.0	99.2	81.6	83.1	99.2	80.5	98.1
14	82.0	81.6	99.2	81.2	81.2	81.6	99.2	81.2	82.8	99.2	80.1	98.1
12	82.0	81.6	99.2	81.2	81.2	81.6	99.2	81.2	82.8	99.2	80.1	98.1
10	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5
8	82.8	82.4	99.6	82.0	82.0	82.4	99.6	82.0	83.5	99.6	80.8	98.5
6	82.4	82.0	99.6	81.6	81.6	82.0	99.6	81.6	83.1	99.6	80.5	98.5

FIG. 2B

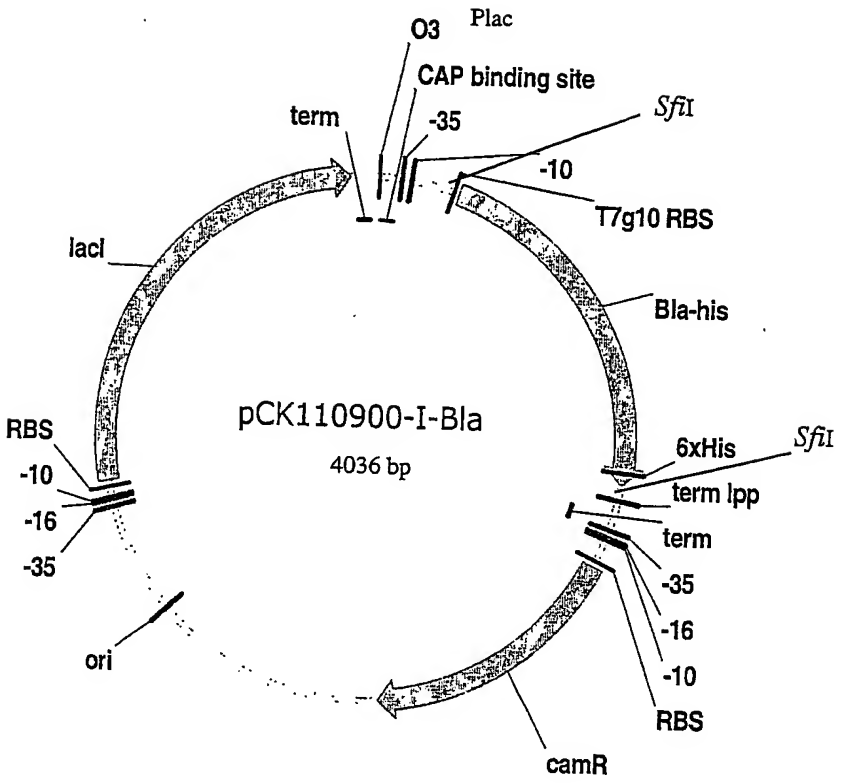

O3
Plac
CAP binding site
term
-35
SfiI
-10
T7g10 RBS
lacI
Bla-his
pCK110900-I-Bla
4036 bp
RBS
-10
-16
-35
6xHis
SfiI
term lpp
term
-35
-16
-10
RBS
ori
camR

FIG. 3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 May 2005 (19.05.2005)

PCT

(10) International Publication Number
WO 2005/045016 A3

(51) International Patent Classification[7]: C12N 9/04, 15/52, 15/63, 5/10

(21) International Application Number: PCT/US2004/026194

(22) International Filing Date: 11 August 2004 (11.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/494,300 11 August 2003 (11.08.2003) US
60/545,657 18 February 2004 (18.02.2004) US

(71) Applicant *(for all designated States except US)*: **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, CA 94063 (US).

(72) Inventors: **KREBBER, Anke**; 3500 Louis Road, Palo Alto, CA 94303 (US). **NEWMAN, Lisa Marie**; 1002 King Street, Redwood City, CA 94061 (US). **DAVIS, S. Christopher**; 104 Winfield Street, San Francisco, CA 94110 (US). **JENNE, Stephane J.**; 821 El Camino Real #202, Burlingame, CA 94010 (US).

(74) Agent: **POCHOPIEN, Donald J.**; MCANDREWS, HELD & MALLOY, LTD., 500 W. Madison Street, Suite 3400, Chicago, IL 60661 (US).

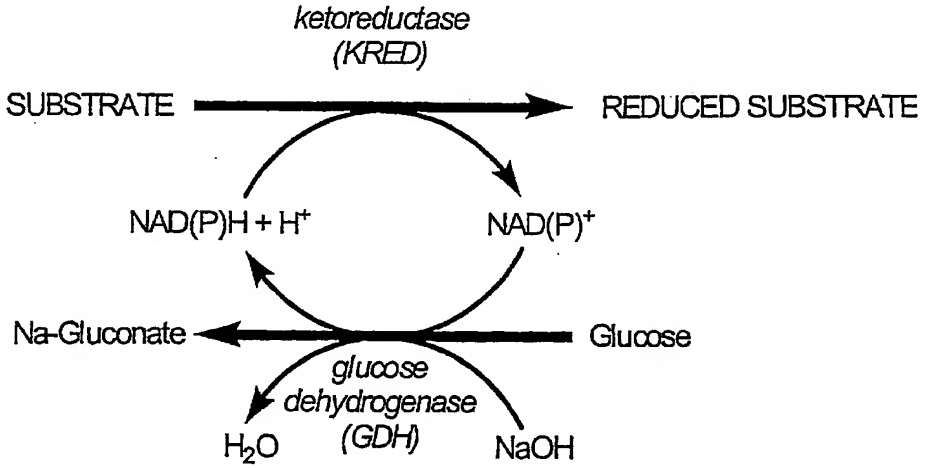
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) Title: IMPROVED GLUCOSE DEHYDROGENASE POLYPEPTIDES AND RELATED POLYNUCLEOTIDES



(57) Abstract: The present invention is directed to glucose dehydrogenase (GDH) polypeptides that have enhanced GDH activity and/or thermostability relative to the backbone wild-type glucose dehydrogenase polypeptide. In addition, the present invention is directed to a plynucleotide that encodes for the GDH polypeptides of the present invention, to nucleic acid sequences comrpising the polynucleotides, to expression vectors comprising the polynucleotides operatively linked to a promoter, to host cells transformed to express the GDH polypeptides, and to a method for producing the GDH polypeptides of the present invention.

WO 2005/045016 A3

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
10 November 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

Inter application No
PCT/US2004/026194

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N9/04 C12N15/52 C12N15/63 C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, PAJ, WPI Data, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LAMPEL K A ET AL: "CHARACTERIZATION OF THE DEVELOPMENTALLY REGULATED BACILLUS-SUBTILIS GLUCOSE DEHYDROGENASE GENE" JOURNAL OF BACTERIOLOGY, vol. 166, no. 1, 1986, pages 238-243, XP002327631 ISSN: 0021-9193 figure 4	1-20

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 May 2005	23. 09. 2005

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Barnas, C

INTERNATIONAL SEARCH REPORT

Inter Application No
PCT/US2004/026194

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MANJON A ET AL: "Increased activity of glucose dehydrogenase co-immobilized with a redox mediator in a bioreactor with electrochemical NAD+ regeneration" BIOTECHNOLOGY LETTERS, vol. 24, no. 15, August 2002 (2002-08), pages 1227-1232, XP002327599 ISSN: 0141-5492 figure 1	1-20
P,X	WO 2004/015132 A (CODEXIS, INC; DAVIS, S., CHRISTOPHER; GRATE, JOHN, H; GRAY, DAVID, R;) 19 February 2004 (2004-02-19) SEQ ID NOs: 62, 66, 68	1-20

INTERNATIONAL SEARCH REPORT

	national application No.
	PCT/US2004/026194

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. type of material

   - [X] a sequence listing
   - [ ] table(s) related to the sequence listing

   b. format of material

   - [ ] in written format
   - [X] in computer readable form

   c. time of filing/furnishing

   - [ ] contained in the international application as filed
   - [X] filed together with the international application in computer readable form
   - [ ] furnished subsequently to this Authority for the purpose of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2004/026194

Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-20

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-20

A polypeptide having at least 1,5 times the glucose dehydrogenase activity of the wild type GDH of SEQ ID NO: 2 being selected from a polypeptide having the amino acid sequence of SEQ ID NOs: 54, 74, 84, 160, 164, 168.

---

2. claim: 21

A polypeptide having glucose dehydrogenase activity and being selected from a polypeptide having the amino acid sequence of SEQ ID NOs: 52, 58, 72.

INTERNATIONAL SEARCH REPORT

Inter application No
PCT/US2004/026194

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004015132 A	19-02-2004	EP 1537222 A2	08-06-2005